

AXIOMS OF ADAPTIVITY FOR SEPARATE MARKING

C. CARSTENSEN^{*†} AND H. RABUS^{*}

Abstract. Mixed finite element methods with flux errors in $H(\text{div})$ -norms and div-least-squares finite element methods require a separate marking strategy in obligatory adaptive mesh-refining. The refinement indicator $\sigma^2(\mathcal{T}, K) = \eta^2(\mathcal{T}, K) + \mu^2(K)$ of a finite element domain K in an admissible triangulation \mathcal{T} consists of some residual-based error estimator $\eta(\mathcal{T}, K)$ with some reduction property under local mesh-refining and some data approximation error $\mu(K)$. Separate marking means either Dörfler marking if $\mu^2(\mathcal{T}) \leq \kappa \eta^2(\mathcal{T})$ or otherwise an optimal data approximation algorithm runs with controlled accuracy as established in [CR11, Rab15].

The axioms are abstract and sufficient conditions on the estimators $\eta(\mathcal{T}, K)$ and data approximation errors $\mu(K)$ for optimal asymptotic convergence rates. The enfolded set of axioms simplifies [CFPP14] for collective marking, treats separate marking established for the first time in an abstract framework, generalizes [CP15] for least-squares schemes, and extends [CR11] to the mixed FEM with flux error control in $H(\text{div})$.

Key words. adaptivity, finite element method, nonstandard finite element method, mixed finite element method, optimal convergence, least-squares finite element method

1. Introduction. The convergence analysis of adaptive finite element methods (AFEMs) with collective marking for some total error estimator (called CAFEM below) is reformulated in an abstract setting in [CFPP14]. Therein four axioms describe elementary properties of the total error estimator that are sufficient for optimal convergence rates. Standard adaptive schemes are based on a total error estimator and collective marking on each level outlined in pseudo code as follows.

```
CAFEM( $\theta, \mathcal{T}_0$ )
  for  $\ell = 0, 1, \dots$ 
    COMPUTE  $\sigma_\ell(K)$  for all  $K \in \mathcal{T}_\ell$ 
     $\mathcal{T}_{\ell+1} := \text{Dörfler\_marking}(\theta, \sigma_\ell(K) : K \in \mathcal{T}_\ell)$ 
```

This paper simplifies the axioms from [CFPP14], also works without the concept of nonlinear approximation classes [BDdV04, Ste07, CKNS08] and so avoids any notion of efficiency. The recent comprehensive a posteriori error analysis in [CPS15] provides an efficient and reliable control in natural norms: the error in the flux in $H(\text{div}, \Omega)$ and the error in the displacements in $L^2(\Omega)$. The focus of this paper is on separate marking (SAFEMs), a modification of the standard AFEM: Dörfler marking is applied if the estimated error dominates the data approximation error, while an optimal data approximation is performed otherwise — outlined in pseudo code as follows.

^{*}Department of Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email cc@math.hu-berlin.de and rabus@math.hu-berlin.de

[†]The work of the first author is partly supported by DFG SPP 1749 *Reliable Simulation Techniques in Solid Mechanics – Development of Non-standard Discretization Methods, Mechanical and Mathematical Analysis*

SAFEM($\theta_A, \kappa, \rho_B, \mathcal{T}_0$)
for $\ell = 0, 1, \dots$
 COMPUTE $\eta_\ell(K), \mu(K)$ for all $K \in \mathcal{T}_\ell$
 if $\mu_\ell^2 := \mu^2(\mathcal{T}_\ell) \leq \kappa \eta_\ell^2 \equiv \kappa \eta_\ell^2(\mathcal{T}_\ell) \eta^2(\mathcal{T}_\ell, \mathcal{T}_\ell)$ // **Case (A)**
 | $\mathcal{T}_{\ell+1} := \text{Dörfler_marking}(\theta_A, \eta_\ell(K) : K \in \mathcal{T}_\ell)$
 else // **Case (B)**
 | $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \text{appx}(\rho_B \mu_\ell^2, \mu(K) : K \in \mathcal{T}_0)$

The algorithm SAFEM combines ideas from [BM08, CR11, Rab15] and distinguishes two Cases (A) and (B), where the refinement is with respect to the dominant refinement indication η_ℓ^2 or μ_ℓ^2 . The refinement in Case (B) depends on the data approximation error and is independent of the discrete solution. This allows for any optimal algorithm for data approximation with respect to the error functional $\mu^2 : K \rightarrow \mathbb{R}$ for $K \subseteq \Omega \subseteq \mathbb{R}^n$, i.e. the output $\mathcal{T}_{\text{Tol}} = \text{appx}(\text{Tol}, \mu(K) : K \in \mathcal{T}_0)$ is expected to satisfy

$$\begin{aligned} \mu^2(\mathcal{T}_{\text{Tol}}) &\leq \text{Tol}, \\ |\mathcal{T}_{\text{Tol}}| - |\mathcal{T}_0| &\leq \Lambda_5 \text{Tol}^{-1/s}. \end{aligned}$$

The analysis for AFEMs based on collective marking as in [CFPP14] is included when $\sigma^2(\mathcal{T}, \bullet) = \eta^2(\mathcal{T}, \bullet) + \mu^2(\mathcal{T}, \bullet)$ replaces $\eta^2(\mathcal{T}, \bullet)$ in Case (A) and the refinement indicator in Case (B) vanishes.

Optimal convergence rates for the estimators follow from axioms (A1)-(A4) generalized from [CFPP14] and (B1)-(B2) for optimal data approximation with quasimonotonicity (QM). The subroutine **appx** in SAFEM can be realized by some Dörfler marking (similar to the algorithm in [BM08]) or by the algorithm APPROX from [BDdV04, BdV04] (applied in [CR11, Rab15]). The flexibility in the data reduction allows applications of SAFEM to problems with data approximation terms that do *not* satisfy an estimator reduction property but quasimonotonicity. Two model examples illustrate this in the present paper: mixed FEM with flux error estimation in $H(\text{div})$ rather than $L^2(\Omega)$ [CR11] and a least-squares FEM problem from [CP15]. Further applications of the present version of the axioms on SAFEM shall appear in the near future [BC, BCS].

The remaining parts of this paper are organised as follows. Section 2 presents more details on SAFEM and guides the reader through the conditions in (A1)-(A4) and (B1)-(B2) for the refinement indicators η and μ and asserts the optimal convergence rate of SAFEM in Theorem 2.1. A collection of remarks follows in Section 3 before Section 4 presents the proofs. Sections 5-6 contain the verification of the axioms for two examples, where separate marking is obligatory for optimal adaptive mesh-refinement. The main novel contribution in Section 5 is the proof of a discrete version (A3) of [CPS15].

The notation $A \lesssim B$ abbreviates $A \leq CB$ for some positive generic constant C , which depends only on the initial triangulation \mathcal{T}_0 and on the universal constants in the axioms; while $A \approx B$ abbreviates $A \lesssim B \lesssim A$. Throughout this paper standard notation of Lebesgue and Sobolev spaces and their norms applies. The modulus sign $|\bullet|$ denotes the Euclidean length as well as the counting measure, e.g., $|\mathcal{M}|$ is the cardinality of \mathcal{M} and equals the number of elements in a triangulation \mathcal{M} (or a subset thereof).

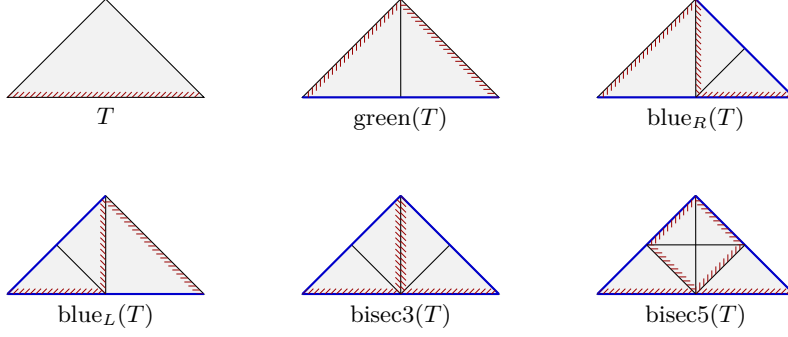


FIG. 2.1. Possible refinements of a triangle depending on the set of marked edges by NVB. Refinement edges are marked red, while marked edges are colored in blue.

2. Axioms and results. The axioms concern general conditions of the estimators η and μ , which play different roles in the adaptive algorithm, and are based on the set \mathbb{T} of admissible triangulations.

2.1. Partitions and admissible triangulations. Let \mathcal{T}_0 be a regular triangulation of the domain Ω into (tagged) n -simplices in \mathbb{R}^n . Any refinement \mathcal{P} from \mathcal{T}_0 by the newest vertex bisection (NVB) of Figure 2.1 is called partition, written $\mathcal{P} \in \mathbb{P}(\mathcal{T}_0) =: \mathbb{P}$. A partition $\mathcal{P} \in \mathbb{P}$, which is a regular triangulation in the sense of Ciarlet, is called admissible, written $\mathcal{P} \in \mathbb{T}(\mathcal{T}_0) =: \mathbb{T}$.

The input of the underlying refinement procedure $\mathcal{T}_{\text{out}} := \text{REFINE}(\mathcal{T}_{\text{in}}, \mathcal{M})$ is an admissible triangulation $\mathcal{T}_{\text{in}} \in \mathbb{T}$ and some subset $\mathcal{M} \subseteq \mathcal{T}_{\text{in}}$ thereof; the output \mathcal{T}_{out} is an admissible triangulation and a one-level refinement of \mathcal{T}_{in} with $\mathcal{M} \subset \mathcal{T}_{\text{in}} \setminus \mathcal{T}_{\text{out}}$ of quasi-minimal cardinality. Conversely, the procedure REFINE specifies the NVB with completion (to avoid hanging nodes etc.) and more details may be found in [Ste08]. NVB is assumed throughout this paper. In particular, given $\mathcal{T}, \mathcal{T}' \in \mathbb{T}$, their overlay $\mathcal{T} \oplus \mathcal{T}' \in \mathbb{T}(\mathcal{T}) \cap \mathbb{T}(\mathcal{T}')$ is the smallest common refinement of \mathcal{T} and \mathcal{T}' .

2.2. Estimators and distance. The axioms are defined in terms of η and μ plus a global distance δ . For any admissible triangulation $\mathcal{T} \in \mathbb{T}$ and any element domain $K \in \mathcal{T}$ let $\eta(\mathcal{T}, K)$ and $\mu(K)$ be a non-negative real number with squares $\eta^2(\mathcal{T}, K)$ and $\mu^2(K)$ and their sums

$$\eta^2(\mathcal{T}, \mathcal{M}) := \sum_{K \in \mathcal{M}} \eta^2(\mathcal{T}, K), \quad \mu^2(\mathcal{M}) := \sum_{K \in \mathcal{M}} \mu^2(K) \quad \text{for any } \mathcal{M} \subseteq \mathcal{T}. \quad (2.1)$$

The distance $\delta(\mathcal{T}, \hat{\mathcal{T}})$ of $\mathcal{T} \in \mathbb{T}$ and its refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ is a non-negative real. The estimators are utilized in the adaptive algorithm and are linked with the distance function in the axioms below. The output of the adaptive algorithm is a sequence $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \dots$ of successive refinements that start with \mathcal{T}_0 and give rise to the abbreviations (with a subindex ℓ to refer to the triangulation as part of the output of SAFEM)

$$\eta_\ell(K) := \eta(\mathcal{T}_\ell, K) \quad \text{for } K \in \mathcal{T}_\ell \quad \text{and} \quad \eta_\ell := \eta(\mathcal{T}_\ell, \mathcal{T}_\ell).$$

The sum $\sigma^2 := \eta^2 + \mu^2$ and their local variants are frequently utilized throughout this paper with $\sigma_\ell^2 := \eta_\ell^2 + \mu_\ell^2$ for $\mu_\ell^2 := \mu^2(\mathcal{T}_\ell) := \sum_{K \in \mathcal{T}_\ell} \mu^2(K)$.

2.3. Adaptive algorithm. In some more details, SAFEM calls SELECT and RE-FINE to realize the Dörfler marking in Case (A) from the introduction; more details on `appx` in Case (B) follow in Subsection 3.3.

SAFEM($\theta_A, \kappa, \rho_B, \mathcal{T}_0$)

Input: Initial coarse triangulation \mathcal{T}_0 , $0 < \theta_A < 1$, $0 < \rho_B < 1$, $0 < \kappa$
for $\ell = 0, 1, \dots$

COMPUTE refinement indicators $\eta_\ell^2(K)$ and $\mu^2(K)$ for all $K \in \mathcal{T}_\ell$
if $\mu_\ell^2 \leq \kappa \eta_\ell^2$ // Case (A)
SELECT a subset $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ of element domains of (almost)
minimal cardinality with
 $\theta_A \eta_\ell^2 \leq \eta_\ell^2(\mathcal{M}_\ell) := \sum_{K \in \mathcal{M}_\ell} \eta_\ell^2(K)$ (2.2)
COMPUTE $\mathcal{T}_{\ell+1} := \text{REFINE}(\mathcal{T}_\ell, \mathcal{M}_\ell)$
else // Case (B)
RUN $\mathcal{T} = \text{appx}(\text{Tol}, \mu(K) : K \in \mathcal{T}_0)$ with $\text{Tol} = \rho_B \mu_\ell^2$
COMPUTE $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \mathcal{T}$

Output: $\mathcal{T}_k, \eta_k, \mu_k, \sigma_k := \sqrt{\eta_k^2 + \mu_k^2}$ for $k = 0, 1, \dots$

The selection of \mathcal{M}_ℓ with almost minimal cardinality means that $|\mathcal{M}_\ell| \lesssim |\mathcal{M}_\ell^*|$, where \mathcal{M}_ℓ^* denotes some set of minimal cardinality with (2.2). The point is that this can be realised in linear CPU time [Ste07].

2.4. Axioms. The universal positive constants $\Lambda_{\text{ref}}, \Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4, \Lambda_6$, and $\hat{\Lambda}_3 \geq 0$ as well as $0 < \rho_2 < 1$ in the axioms (A1)-(A4), (B2), and (QM) below solely depend on \mathbb{T} (whence merely on \mathcal{T}_0); the parameters $s > 0$ and Λ_5 in (B1) also depend on the algorithm `appx` and the optimal data approximation rate.

The axioms (A1)-(A3) and (B2) concern an arbitrary triangulation $\mathcal{T} \in \mathbb{T}$ and any refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ of it, while (A4) solely concerns the outcome of SAFEM. Recall the sum conventions for $\eta(\mathcal{T}, \mathcal{M})$ and $\mu(\mathcal{T})$ in Subsection 2.2.

(A1) Stability. $\forall \mathcal{T} \in \mathbb{T} \forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$

$$\left| \eta(\hat{\mathcal{T}}, \mathcal{T} \cap \hat{\mathcal{T}}) - \eta(\mathcal{T}, \mathcal{T} \cap \hat{\mathcal{T}}) \right| \leq \Lambda_1 \delta(\mathcal{T}, \hat{\mathcal{T}}). \quad (\text{A1})$$

(A2) Reduction. $\forall \mathcal{T} \in \mathbb{T} \forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$

$$\eta(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T}) \leq \rho_2 \eta(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}}) + \Lambda_2 \delta(\mathcal{T}, \hat{\mathcal{T}}). \quad (\text{A2})$$

(A3) Discrete Reliability. $\forall \mathcal{T} \in \mathbb{T} \forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T}) \exists \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) \subseteq \mathcal{T}$ with $\mathcal{T} \setminus \hat{\mathcal{T}} \subseteq \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}})$,

$$\begin{aligned} \left| \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) \right| &\leq \Lambda_{\text{ref}} \left| \mathcal{T} \setminus \hat{\mathcal{T}} \right| \quad \text{and} \\ \delta^2(\mathcal{T}, \hat{\mathcal{T}}) &\leq \Lambda_3 \left(\eta^2(\mathcal{T}, \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}})) + \mu^2(\mathcal{T}) \right) + \hat{\Lambda}_3 \eta^2(\hat{\mathcal{T}}). \end{aligned} \quad (\text{A3})$$

(A4) Quasiorthogonality of discrete solutions. $\forall \ell \in \mathbb{N}_0$

$$\sum_{k=\ell}^{\infty} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) \leq \Lambda_4 \sigma_\ell^2. \quad (\text{A4})$$

(B1) Rate s data approximation. $\forall \text{Tol} > 0$, $\mathcal{T}_{\text{Tol}} := \text{appx}(\text{Tol}, \mu(K) : K \in \mathcal{T}_0) \in \mathbb{T}$ satisfies

$$|\mathcal{T}_{\text{Tol}}| - |\mathcal{T}_0| \leq \Lambda_5 \text{Tol}^{-1/(2s)} \quad \text{and} \quad \mu^2(\mathcal{T}_{\text{Tol}}) \leq \text{Tol}. \quad (\text{B1})$$

(B2) Quasimonotonicity of μ . $\forall \mathcal{T} \in \mathbb{T} \forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T}) \quad \mu(\hat{\mathcal{T}}) \leq \Lambda_6 \mu(\mathcal{T})$.

Theorem 3.2 below asserts that the aforementioned axioms imply quasimonotonicity of σ for small values of $\hat{\Lambda}_3$, while this axiom (QM) stands on its own in the example of Section 6.

(QM) Quasimonotonicity of σ . $\forall \mathcal{T} \in \mathbb{T} \forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T}) \quad \sigma(\hat{\mathcal{T}}) \leq \Lambda_7 \sigma(\mathcal{T})$.

2.5. Optimal convergence rates. The axioms (A1)-(A4), (B1)-(B2), and (QM) ensure quasioptimality of SAFEM for sufficiently small parameters θ_A and κ as stated in Theorem 2.1 below. Recall that $\sigma^2 := \eta^2 + \mu^2$ and set

$$\sigma^2(\mathcal{T}) \equiv \sigma(\mathcal{T})^2 := \sigma^2(\mathcal{T}, \mathcal{T}) := \sum_{K \in \mathcal{T}} \sigma^2(\mathcal{T}, K) \text{ for } \mathcal{T} \in \mathbb{T} \text{ and } \sigma_\ell := \sigma(\mathcal{T}_\ell).$$

For any $N \in \mathbb{N}_0$, the comparison with the optimal rates concern the optimal value

$$\min \sigma(\mathbb{T}(N)) := \min \{ \sigma(\mathcal{T}) : \mathcal{T} \in \mathbb{T}(N) \}$$

of all admissible triangulations

$$\mathbb{T}(N) := \{ \mathcal{T} \in \mathbb{T} : |\mathcal{T}| \leq |\mathcal{T}_0| + N \}$$

of cardinality $|\mathcal{T}| \leq |\mathcal{T}_0| + N$ with at most N extra cells.

THEOREM 2.1 (Quasioptimality). *Suppose (A1)-(A4) and (B1)-(B2). (a) The strict inequality $(\Lambda_1^2 + \Lambda_2^2)\hat{\Lambda}_3 < 1$ implies (QM) with Λ_7 depending on $\Lambda_1, \Lambda_2, \Lambda_3, \hat{\Lambda}_3$, and Λ_6 . (b) The axiom (QM) leads to the existence of some $\kappa_0 > 0$, which is $+\infty$ if $\Lambda_6 = 1$, such that any choice of κ, θ_A , and ρ_B with*

$$0 < \kappa < \kappa_1 := \min \{ \kappa_0, \Lambda_1^{-2} \Lambda_3^{-1} \}, \quad 0 < \theta_A < \theta_0 := (1 - \kappa \Lambda_1^2 \Lambda_3) / (1 + \Lambda_1^2 \Lambda_3),$$

and $0 < \rho_B < 1$ implies the following. The output $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ and $(\sigma_\ell)_{\ell \in \mathbb{N}_0}$ of SAFEM satisfy the equivalence

$$\Lambda_5^s + \sup_{\ell \in \mathbb{N}_0} (1 + |\mathcal{T}_\ell| - |\mathcal{T}_0|)^s \sigma_\ell \approx \Lambda_5^s + \sup_{N \in \mathbb{N}_0} (1 + N)^s \min \sigma(\mathbb{T}(N)). \quad (2.3)$$

In particular, the left-hand side of the equivalence (2.3) is smaller than infinity if the right-hand is and vice versa. The quotient is bounded below and from above by the equivalence constants, which depend on $\Lambda_{\text{ref}}, \Lambda_1, \Lambda_2, \Lambda_3, \hat{\Lambda}_3, \Lambda_4, \Lambda_6, \rho_B, \rho_2, \theta_A, \kappa$, and s but not on Λ_5 .

The (possibly unknown) parameter s is not utilized in SAFEM. The axioms (B1)-(B2) specify *sufficient* conditions for optimal convergence, where the parameter $s > 0$ is arbitrary and may refer to a related nonlinear approximation class.

3. Remarks.

3.1. Weak form of (A4). The axiom (A4) can be weakened with some parameter $\varepsilon > 0$, which vanishes in (A4) \equiv (A4₀).

(A4_ε) Quasiorthogonality with $\varepsilon > 0$. $\exists \varepsilon > 0 \exists 0 < \Lambda_{4(\varepsilon)} < \infty \forall \ell, m \in \mathbb{N}_0$

$$\sum_{k=\ell}^{\ell+m} \delta_{k,k+1}^2 \leq \Lambda_{4(\varepsilon)} \sigma_\ell^2 + \varepsilon \sum_{k=\ell}^{\ell+m} \sigma_k^2. \quad (\text{A4}_\varepsilon)$$

The axiom (A4_ε) implies (A4_{ε'}) for all $0 \leq \varepsilon < \varepsilon'$ with the same constant $\Lambda_{4(\varepsilon)} = \Lambda_{4(\varepsilon')}$, and (A4) is (A4₀), i.e. (A4_ε) for $\varepsilon = 0$. Conversely, as $\varepsilon \searrow 0$ it may be expected that $\Lambda_{4(\varepsilon)} \rightarrow \infty$. In the presence of (A1)-(A2), this is not the case. In fact, (A1)-(A2) and (A4_ε) imply (A4) for sufficiently small $\varepsilon > 0$.

THEOREM 3.1 ((A4_ε) \Rightarrow (A4)). *Let θ_A be the parameter of SAFEM and $0 < \rho_{12} < 1$ the reduction factor for the total error estimator with constant $0 < \Lambda_{12} < \infty$ in Theorem 4.1 below and let $0 \leq \varepsilon < (1 - \rho_{12})/\Lambda_{12}$. Then (A1)-(A2) and (A4_ε) imply (A4) with $\Lambda_4 := \Lambda_{4(\varepsilon)} + \varepsilon(1 + \Lambda_{12}\Lambda_{4(\varepsilon)})/(1 - \rho_{12} - \varepsilon\Lambda_{12})$.*

This has first been observed in [CFPP14] for CAFEM and is proved in Subsection 4.2 for completeness and applied below in Theorem 5.1.

3.2. Quasimonotonicity. The axiom (B2) explicitly ensures the quasimonotonicity of μ and (QM) follows with $\Lambda_7 := \sqrt{\Lambda_6^2 + \Lambda_8^2}$ from the subsequent theorem: $\widehat{\Lambda}_3 < 1/(\Lambda_1^2 + \Lambda_2^2)$ is sufficient for (QM).

THEOREM 3.2 (Quasimonotonicity). *Suppose (A1)-(A3) and $\widehat{M} := (\Lambda_1^2 + \Lambda_2^2)\widehat{\Lambda}_3 < 1$. Set $M := (\Lambda_1^2 + \Lambda_2^2)\Lambda_3$ and*

$$\Lambda_8 := \frac{1 + M(1 - \widehat{M}) + \widehat{M} + 2\sqrt{M(1 - \widehat{M}) + \widehat{M}}}{(1 - \widehat{M})^2}.$$

Then, any $\mathcal{T} \in \mathbb{T}$ and $\widehat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ satisfy

$$\eta(\widehat{\mathcal{T}}) \leq \Lambda_8 \sigma(\mathcal{T}). \quad (3.1)$$

Proof. Given $\lambda := (\sqrt{M + \widehat{M} - M\widehat{M}} - \widehat{M})/(M + \widehat{M}) < 1/\widehat{M} - 1$, recall the following implication of the axioms (A1)-(A3), namely

$$\begin{aligned} \eta^2(\widehat{\mathcal{T}}, \widehat{\mathcal{T}} \cap \mathcal{T}) &\leq (1 + 1/\lambda) \eta^2(\mathcal{T}, \widehat{\mathcal{T}} \cap \mathcal{T}) + (1 + \lambda) \Lambda_1^2 \delta^2(\mathcal{T}, \widehat{\mathcal{T}}), \\ \eta^2(\widehat{\mathcal{T}}, \widehat{\mathcal{T}} \setminus \mathcal{T}) &\leq (1 + 1/\lambda) \rho_2^2 \eta^2(\mathcal{T}, \mathcal{T} \setminus \widehat{\mathcal{T}}) + (1 + \lambda) \Lambda_2^2 \delta^2(\mathcal{T}, \widehat{\mathcal{T}}), \\ \delta^2(\mathcal{T}, \widehat{\mathcal{T}}) &\leq \Lambda_3 \sigma^2(\mathcal{T}) + \widehat{\Lambda}_3 \eta^2(\widehat{\mathcal{T}}). \end{aligned}$$

Those inequalities plus the split $\eta^2(\widehat{\mathcal{T}}) = \eta^2(\widehat{\mathcal{T}}, \widehat{\mathcal{T}} \cap \mathcal{T}) + \eta^2(\widehat{\mathcal{T}}, \widehat{\mathcal{T}} \setminus \mathcal{T})$ verify

$$\eta^2(\widehat{\mathcal{T}}) \leq (1 + 1/\lambda) \eta^2(\mathcal{T}) + (1 + \lambda) (\Lambda_1^2 + \Lambda_2^2) (\Lambda_3 \sigma^2(\mathcal{T}) + \widehat{\Lambda}_3 \eta^2(\widehat{\mathcal{T}})). \quad \square$$

3.3. Optimal data approximation with APPROX. Case (B) of SAFEM runs a data approximation algorithm `appx`(Tol, $\mu(K) : K \in \mathcal{T}_0$) with output in \mathbb{T} . The data approximation algorithm APPROX [BDdV04, BdV04] is based on the refinement of partitions and has been established for separate marking algorithms in [CR11, Rab15] and is one possible realisation of `appx` in SAFEM.

Let $\widehat{\mathcal{P}}$ be some NVB refinement of $\mathcal{P} \in \mathbb{P}$. Let $K \in \mathcal{P}$ and $\widehat{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, then the refinement of K in $\widehat{\mathcal{P}}$ is the set $\widehat{\mathcal{P}}(K) := \{T \in \widehat{\mathcal{P}} \mid T \subseteq K\}$ in the following.

(SA) Sub-additivity. $\exists \Lambda_6 < \infty \forall \mathcal{P} \in \mathbb{P} \forall \hat{\mathcal{P}} \in \mathbb{P}(\mathcal{P}) \forall \mathcal{M} \subseteq \mathcal{P}$

$$\mu^2(\hat{\mathcal{P}}(\mathcal{M})) := \sum_{K \in \mathcal{M}} \sum_{T \in \hat{\mathcal{P}}(K)} \mu^2(T) \leq \Lambda_6 \mu^2(\mathcal{M}). \quad (\text{SA})$$

Note, that the notation of the data approximation term μ is a straight forward extension of its definition in (2.1) for admissible triangulations to partitions.

The algorithm APPROX is outlined in the following with input tolerance $\text{Tol}' := \text{Tol} / \Lambda_6 = \rho_B \mu_\ell / \Lambda_6$ and the values $\mu(K)$ on the coarse triangulation \mathcal{T}_0 .

APPROX ($\text{Tol}', \mu(K) : K \in \mathcal{T}_0$)

COMPUTE $\tilde{\mu}^2(T) = \mu^2(T)$ for all $T \in \mathcal{T}_0$ and set $\mu^2(\mathcal{T}_0) := \sum_{T \in \mathcal{T}_0} \mu^2(T)$

SET $\mathcal{P} = \mathcal{T}_0$

while $\mu^2(\mathcal{P}) > \text{Tol}'$ **do**

 COMPUTE $\tilde{\mu}^2(T)$ for all $T \in \mathcal{P}$, set $\tilde{\mu}_{\max}^2 := \max_{T \in \mathcal{P}} \tilde{\mu}^2(T)$

 SELECT a subset $\mathcal{M} := \{T \in \mathcal{P} \mid \tilde{\mu}^2(T) = \tilde{\mu}_{\max}^2\} \subseteq \mathcal{P}$

 COMPUTE $\overline{\mathcal{P}} := \text{bisec}(\mathcal{P}, \mathcal{M})$

COMPUTE $\mathcal{T}_{\text{Tol}} := \text{completion}(\mathcal{P}) \in \mathbb{T}$

REMARK 1. (a) Algorithm APPROX is based on a modified error functional $\tilde{\mu}$ initiated by $\tilde{\mu}(K) := \mu(K)$ for all $K \in \mathcal{T}_0$. Given $\tilde{\mu}(K)$ for a triangle $K = K_1 \cup K_2$ bisected into sub-triangles K_1 and K_2 , let

$$\tilde{\mu}(K_j) := \frac{\tilde{\mu}(K)(\mu(K_1) + \mu(K_2))}{\mu(K) + \tilde{\mu}(K)} \quad \text{for } j = 1, 2. \quad (3.2)$$

(b) Notice that the partitions \mathcal{P} in the while-loop in APPROX are not regular in general and the final completion step may be realized with successive calls of REFINE.

(c) The implementation of APPROX may store the partition \mathcal{P} and the values $\tilde{\mu}(K)$ for all element domains $K \in \mathcal{P}$ at the end of the while loop to keep the successive calls of APPROX for various decreasing tolerances Tol' efficient.

THEOREM 3.3 ([BdV04, BDdV04]). (SA) in APPROX implies (B1)-(B2) with rate-s-optimality in the sense that

$$M(s, \mu) := \sup_{N \in \mathbb{N}_0} (1 + N)^s \min \mu(\mathbb{T}(N)) \approx \Lambda_5^s \quad (3.3)$$

holds for all $s > 0$ (and $M(s, \mu) < \infty$ if and only if $\Lambda_5 < \infty$).

Proof. This follows from near optimality proven in [BdV04, Theorem 6.1] and [BDdV04, Lemma 4.4]. \square

3.4. Collective Dörfler marking is optimal for $\|h_\ell f\|_{L^2(\Omega)}$. Given $f \in L^2(\Omega)$ in the polyhedral domain $\Omega \subseteq \mathbb{R}^n$ partitioned into the regular triangulation \mathcal{T}_0 , set $\eta(\mathcal{T}_\ell, K) := |K|^{2/n} |f|_{L^2(K)}$ for all $K \in \mathcal{T}_\ell$. Let $\eta_\ell = \eta(\mathcal{T}_\ell, \mathcal{T}_\ell)$. Then, (A1)-(A4) are satisfied with appropriate weight functions $h_\mathcal{T}$ (resp. $h_{\hat{\mathcal{T}}}$) of mesh-sizes in \mathcal{T} (resp. $\hat{\mathcal{T}}$)

$$\delta(\mathcal{T}, \hat{\mathcal{T}}) := \|(h_\mathcal{T} - h_{\hat{\mathcal{T}}})f\|_{L^2(\Omega)}.$$

Hence CAFEM with collective Dörfler marking implies optimal data approximation for this particular data error term with a mesh-size weight $h_\mathcal{T}$. This is in agreement with the well-established fact that first-order conforming and nonconforming finite element methods do not need a data reduction with SAFEM.

4. Proofs. The abbreviation $\delta_{\ell,\ell+1} := \delta(\mathcal{T}_\ell, \mathcal{T}_{\ell+1})$ applies throughout this section.

4.1. Estimator reduction. The constant $\Lambda_6 \geq 1$ in the following theorem leads to κ_0 set to $+\infty$ for $\Lambda_6 = 1$; $\kappa_0 = \infty$ and $\Lambda_6 = 1$ hold in all the examples of this paper.

THEOREM 4.1 ((A12) reduction). *Suppose (A1)-(A2) and parameters $0 < \theta_A \leq 1$, $0 < \kappa$, and $0 < \rho_B < 1/\Lambda_6$ from SAFEM. Any choice of γ and λ with*

$$0 < \gamma < \rho_2^{-2} - 1 \text{ and } 0 < \lambda < \min \left\{ (1 - (1 + \gamma)\rho_2^2) \frac{\theta_A}{1 - \theta_A}, \kappa(1 - \rho_B) \right\} \quad (4.1)$$

lead to constants

$$0 < \Lambda_{12} := (1 + 1/\lambda)\Lambda_1^2 + (1 + 1/\gamma)\Lambda_2^2 < \infty, \quad (4.2)$$

$$0 < \rho_A := (1 + \lambda)(1 - \theta_A) + (1 + \gamma)\rho_2^2\theta_A < 1, \quad (4.3)$$

$$0 < \kappa_0 := (1 - \rho_A)/(\Lambda_6 - 1) \text{ (with } \kappa_0 := +\infty \text{ if } \Lambda_6 = 1), \quad (4.4)$$

$$0 < \rho_{12} := \max \{ \rho_A + \kappa\Lambda_6, 1 + \lambda + \kappa\rho_B \} / (1 + \kappa) \leq 1. \quad (4.5)$$

Moreover, $0 < \kappa < \kappa_0$ implies $\rho_{12} < 1$ and

$$\sigma_{\ell+1}^2 \leq \rho_{12}\sigma_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 \quad \text{for all } \ell \in \mathbb{N}_0 \quad (\text{A12})$$

for the output σ_ℓ^2 of SAFEM.

Proof For γ and λ as in (4.1), the axioms (A1)-(A2) imply

$$\begin{aligned} \eta_{\ell+1}^2(\mathcal{T}_{\ell+1} \cap \mathcal{T}_\ell) &\leq (1 + \lambda)\eta_\ell^2(\mathcal{T}_{\ell+1} \cap \mathcal{T}_\ell) + (1 + 1/\lambda)\Lambda_1^2\delta_{\ell,\ell+1}^2, \\ \eta_{\ell+1}^2(\mathcal{T}_{\ell+1} \setminus \mathcal{T}_\ell) &\leq (1 + \gamma)\rho_2^2\eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}) + (1 + 1/\gamma)\Lambda_2^2\delta_{\ell,\ell+1}^2. \end{aligned}$$

The sum of those two inequalities leads to

$$\eta_{\ell+1}^2 \leq (1 + \lambda)\eta_\ell^2 + ((1 + \gamma)\rho_2^2 - (1 + \lambda))\eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}) + \Lambda_{12}\delta_{\ell,\ell+1}^2. \quad (4.6)$$

The restrictions on λ and γ ensure $(1 + \gamma)\rho_2^2 < 1 < 1 + \lambda$. Thus, in general,

$$\eta_{\ell+1}^2 \leq (1 + \lambda)\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

In Case (A) on the level ℓ , when Dörfler's marking ensures $\theta_A\eta_\ell^2 \leq \eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1})$, this and (4.6) leads to an improvement of the last estimate, namely

$$\eta_{\ell+1}^2 \leq ((1 + \lambda)(1 - \theta_A) + (1 + \gamma)\rho_2^2\theta_A)\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 = \rho_A\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

The restrictions on λ and γ reveal $\rho_A < 1$. Altogether, let

$$R_\ell := \begin{cases} \rho_A & \text{in Case (A) on level } \ell, \\ 1 + \lambda & \text{in Case (B) on level } \ell. \end{cases} \quad (4.7)$$

Then, the output of SAFEM satisfies

$$\eta_{\ell+1}^2 \leq R_\ell\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 \quad \text{for all } \ell \in \mathbb{N}_0. \quad (4.8)$$

In Case (A) on any level ℓ with $R_\ell = \rho_A$ from (4.3) and Λ_{12} from (4.2), it also holds $\mu_{\ell+1}^2 \leq \Lambda_6 \mu_\ell^2$, and $\mu_\ell^2 \leq \kappa \eta_\ell^2$. Since $\alpha := (\Lambda_6 - \rho_A)/(\kappa + 1) > 0$, this and (4.8) lead to

$$\sigma_{\ell+1}^2 \leq (\rho_A + \alpha \kappa) \eta_\ell^2 + (\Lambda_6 - \alpha) \mu_\ell^2 + \Lambda_{12} \delta_{\ell,\ell+1}^2 = \frac{\rho_A + \kappa \Lambda_6}{1 + \kappa} \sigma_\ell^2 + \Lambda_{12} \delta_{\ell,\ell+1}^2.$$

In Case (B) on the level ℓ with $R_\ell = 1 + \lambda$, it holds $\mu_{\ell+1}^2 \leq \rho_B \mu_\ell^2$, and $\kappa \eta_\ell^2 < \mu_\ell^2$. Since $\beta := \kappa(1 + \lambda - \rho_B)/(1 + \kappa) > 0$, this and (4.8) lead to

$$\sigma_{\ell+1}^2 < (1 + \lambda - \beta) \eta_\ell^2 + (\rho_B + \beta/\kappa) \mu_\ell^2 + \Lambda_{12} \delta_{\ell,\ell+1}^2 = \frac{1 + \kappa \rho_B + \lambda}{1 + \kappa} \sigma_\ell^2 + \Lambda_{12} \delta_{\ell,\ell+1}^2.$$

This proves the total error estimator reduction (A12) with ρ_{12} from (4.5). \square

4.2. Convergence. The plain convergence follows from the estimator reduction (A12) plus quasiorthogonality (A4).

THEOREM 4.2. *Suppose $0 < \theta_A \leq 1$, $0 < \kappa$, $0 < \rho_B < 1$, suppose (A4) and (A12) with constants $0 < \rho_{12} < 1$ and $0 < \Lambda_{12} < \infty$. Then $\Lambda := (1 + \Lambda_{12} \Lambda_4)/(1 - \rho_{12})$, $q := \Lambda/(1 + \Lambda) < 1$, and the output of SAFEM satisfy the following assertions (a)-(c).*

- (a) (Plain convergence) $\forall \ell, m \in \mathbb{N}_0 \quad \sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq \Lambda \sigma_\ell^2$.
- (b) (R-linear convergence on each level) $\forall \ell, m \in \mathbb{N}_0 \quad \sigma_{\ell+m}^2 \leq \frac{q^m}{1-q} \sigma_\ell^2$.
- (c) (Reciprocal sum) $\forall s > 0 \quad \forall \ell \in \mathbb{N} \quad \sum_{k=0}^{\ell-1} \sigma_k^{-1/s} \leq \frac{q^{1/(2s)} \sigma_\ell^{-1/s}}{(1-q)^{1/(2s)} (1-q^{1/(2s)})}$.

Proof of (a). For all $\ell, m \in \mathbb{N}_0$, (A12) implies

$$\sum_{k=\ell}^{\ell+m} \sigma_k^2 = \sigma_\ell^2 + \sum_{k=\ell+1}^{\ell+m} \sigma_k^2 \leq \sigma_\ell^2 + \rho_{12} \sum_{k=\ell}^{\ell+m} \sigma_k^2 + \Lambda_{12} \sum_{k=\ell}^{\ell+m} \delta_{k,k+1}^2. \quad (4.9)$$

This plus (A4) verify

$$(1 - \rho_{12}) \sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq \sigma_\ell^2 + \Lambda_{12} \Lambda_4 \sigma_\ell^2.$$

This proves (a) with the asserted constant Λ . \square

Proof of Theorem 3.1. The same argument as in the proof of (a) before show that (A12) and (A4 $_\varepsilon$) imply (A4) for small ε . In fact, (4.9) and (A4 $_\varepsilon$) show

$$(1 - \rho_{12}) \sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq \sigma_\ell^2 + \Lambda_{12} \left(\Lambda_{4(\varepsilon)} \sigma_\ell^2 + \varepsilon \sum_{k=\ell}^{\ell+m} \sigma_k^2 \right).$$

In other words

$$(1 - \rho_{12} - \varepsilon \Lambda_{12}) \sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq (1 + \Lambda_{12} \Lambda_{4(\varepsilon)}) \sigma_\ell^2.$$

This plus (A4_ε) lead to (A4) with $\Lambda_4 := \Lambda_{4(\varepsilon)} + \varepsilon(1 + \Lambda_{12}\Lambda_{4(\varepsilon)})/(1 - \rho_{12} - \varepsilon\Lambda_{12})$

$$\sum_{k=\ell}^{\ell+m} \delta_{k,k+1}^2 \leq \Lambda_{4(\varepsilon)} \sigma_\ell^2 + \varepsilon \sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq \Lambda_4 \sigma_\ell^2. \quad \square$$

Proof of Theorem 4.2.b. The assertion (a) implies the convergence of the series

$$\xi_{\ell+1}^2 := \sum_{k=\ell+1}^{\infty} \sigma_k^2 \leq \Lambda \sigma_\ell^2 < \infty.$$

The addition of $\Lambda \xi_{\ell+1}^2$ to the previous inequality results in

$$(\Lambda + 1) \xi_{\ell+1}^2 \leq \Lambda \xi_\ell^2, \text{ hence } \xi_{\ell+1}^2 \leq q \xi_\ell^2. \quad (4.10)$$

The successive application of the previous contraction (4.10) shows

$$\sigma_{\ell+m}^2 \leq \xi_{\ell+m}^2 \leq q^m \xi_\ell^2 = q^m (\sigma_\ell^2 + \xi_{\ell+1}^2) \leq q^m (1 + \Lambda) \sigma_\ell^2. \quad \square$$

Proof of Theorem 4.2.c. The R-linear convergence of (b) leads to

$$\sigma_k^{-1/s} \leq \frac{q^{(\ell-k)/(2s)}}{(1-q)^{1/(2s)}} \sigma_\ell^{-1/s} \quad \text{for all } 0 \leq k < \ell.$$

This proves

$$\sum_{k=0}^{\ell-1} \sigma_k^{-1/s} \leq \frac{\sigma_\ell^{-1/s}}{(1-q)^{1/(2s)}} \sum_{k=0}^{\ell-1} \left(q^{1/(2s)} \right)^{\ell-k} \leq \frac{\sigma_\ell^{-1/s} q^{1/(2s)}}{(1-q)^{1/(2s)} (1 - q^{1/(2s)})}. \quad \square$$

LEMMA 4.3 (Comparison). *Suppose (A1)-(A4), (B1)-(B2) with $0 < s < \infty$, (QM), $0 < q < 1$ from Theorem 4.2.b, and let $0 < \xi < 1$ and $0 < \nu < \infty$; let*

$$M := M(s, \sigma) := \sup_{N \in \mathbb{N}_0} (N+1)^s \min \sigma(\mathbb{T}(N)) < \infty, \quad (4.11)$$

similar to the definition of $M(s, \mu)$ in (3.3). Then for any level $\ell \in \mathbb{N}_0$ of SAFEM with a triangulation \mathcal{T}_ℓ , there exists a refinement $\hat{\mathcal{T}}_\ell \in \mathbb{T}(\mathcal{T}_\ell)$ with (a)-(c).

- (a) $\sigma(\hat{\mathcal{T}}_\ell) \leq \xi \sigma_\ell$;
- (b) $\sqrt{1 - q\xi} \sigma_\ell \left| \mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell \right|^s \leq \Lambda_7 M$;
- (c) $\left(1 - \xi^2(1 + \nu + (1 + 1/\nu)\Lambda_1^2 \hat{\Lambda}_3) \right) \eta_\ell^2$
 $\leq (1 + (1 + 1/\nu)\Lambda_1^2 \Lambda_3) \eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell))$
 $+ \left((1 + \nu)\xi^2 + (1 + 1/\nu)\Lambda_1^2(\Lambda_3 + \hat{\Lambda}_3 \xi^2) \right) \mu_\ell^2.$

Proof. Two pathological situations are excluded in the beginning of the proof. First, if $\sigma_\ell = 0$, then $\hat{\mathcal{T}}_\ell = \mathcal{T}_\ell$ satisfies the assumptions (a)-(c). Second, Theorem 4.2 guarantees convergence of some sequence of triangulations and (QM) implies that this holds for uniform refinements as well. Hence there exists a refinement $\hat{\mathcal{T}}_\ell$ of \mathcal{T}_ℓ with (a) and $\hat{\mathcal{T}}_\ell \cap \mathcal{T}_\ell = \emptyset$. The latter implies (c) even in case $M \equiv M(s, \sigma) = \infty$ when (b) is obvious.

Throughout the remaining parts of the proof, it is therefore assumed that $M < \infty$ and $\sigma_\ell > 0$. Then (QM) implies $0 < \sigma_0 \leq M < \infty$.

1. *Setup.* Let $N_\ell \in \mathbb{N}_0$ be minimal with

$$(N_\ell + 1)^{-s} \leq \frac{\xi \sqrt{1-q}}{\Lambda_7 M} \sigma_\ell. \quad (4.12)$$

The quasimonotonicity (QM) followed by the definition of $M := M(s, \sigma) < \infty$ in (4.11) and $0 < q < 1, 0 < \xi < 1$ lead to

$$\frac{\xi \sqrt{1-q}}{\Lambda_7} \sigma_\ell \leq \xi \sqrt{1-q} \sigma_0 \leq \xi \sqrt{1-q} M < M.$$

Hence, $(N_\ell + 1)^{-s} < 1$ and so $N_\ell \geq 1$. Since $N_\ell \in \mathbb{N}$ is minimal with (4.12),

$$0 < (N_\ell + 1)^{-s} \leq \frac{\xi \sqrt{1-q}}{\Lambda_7 M} \sigma_\ell < N_\ell^{-s}.$$

This implies

$$N_\ell^s < \frac{\Lambda_7 M}{\xi \sqrt{1-q}} \sigma_\ell^{-1}. \quad (4.13)$$

2. *Design of $\hat{\mathcal{T}}_\ell$.* The definition of $M < \infty$ yields the existence of some optimal $\tilde{\mathcal{T}}_\ell \in \mathbb{T}(N_\ell)$ with

$$(N_\ell + 1)^s \sigma(\tilde{\mathcal{T}}_\ell) \leq M. \quad (4.14)$$

The overlay triangulation $\hat{\mathcal{T}}_\ell := \mathcal{T}_\ell \oplus \tilde{\mathcal{T}}_\ell$ [CKNS08, Ste07] satisfies

$$|\hat{\mathcal{T}}_\ell| + |\mathcal{T}_0| \leq |\mathcal{T}_\ell| + |\tilde{\mathcal{T}}_\ell|. \quad (4.15)$$

3. *Proof of (a).* The quasimonotonicity (QM) followed by (4.14) and (4.12) shows

$$\sigma(\hat{\mathcal{T}}_\ell) \leq \Lambda_7 \sigma(\tilde{\mathcal{T}}_\ell) \leq \Lambda_7 M (N_\ell + 1)^{-s} \leq \xi \sigma_\ell \sqrt{1-q} < \xi \sigma_\ell. \quad \square$$

4. *Proof of (b).* The definition of $\tilde{\mathcal{T}}_\ell$, the overlay estimate in (4.15), and the upper bound for N_ℓ in (4.13) lead to

$$|\mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell| \leq |\hat{\mathcal{T}}_\ell| - |\mathcal{T}_\ell| \leq |\tilde{\mathcal{T}}_\ell| - |\mathcal{T}_0| \leq N_\ell \leq \left(\frac{\Lambda_7 M}{\xi \sigma_\ell \sqrt{1-q}} \right)^{1/s}. \quad \square$$

5. *Proof of (c).* For any $0 < \nu < \infty, 0 < \xi < 1$, (A1) and (A3) result in

$$\begin{aligned} \eta_\ell^2(\mathcal{T}_\ell \cap \hat{\mathcal{T}}_\ell) &\leq (1 + \nu) \eta^2(\hat{\mathcal{T}}_\ell, \hat{\mathcal{T}}_\ell \cap \mathcal{T}_\ell) + (1 + 1/\nu) \Lambda_1^2 \delta^2(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell) \\ &\leq \left(1 + \nu + (1 + 1/\nu) \Lambda_1^2 \hat{\Lambda}_3 \right) \eta^2(\hat{\mathcal{T}}_\ell) \\ &\quad + (1 + 1/\nu) \Lambda_1^2 \Lambda_3 \left(\eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)) + \mu_\ell^2 \right). \end{aligned}$$

This, (a), and $\mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell \subseteq \mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)$ result in

$$\begin{aligned} \eta_\ell^2 &= \eta_\ell^2(\mathcal{T}_\ell \cap \hat{\mathcal{T}}_\ell) + \eta_\ell^2(\mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell) \\ &\leq \left(1 + \nu + (1 + 1/\nu) \Lambda_1^2 \hat{\Lambda}_3 \right) \xi^2 \sigma_\ell^2 + (1 + (1 + 1/\nu) \Lambda_1^2 \Lambda_3) \eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)) \\ &\quad + (1 + 1/\nu) \Lambda_1^2 \Lambda_3 \mu_\ell^2. \end{aligned}$$

Some rearrangements with $\sigma_\ell^2 = \eta_\ell^2 + \mu_\ell^2$ prove (c). \square

4.3. Proof of Theorem 2.1. *Proof of Theorem 2.1.a.* This is a consequence of Theorem 3.2 plus (B2). \square

Proof of “ \lesssim ” in (2.3) of Theorem 2.1.b. Since $\theta_A < \theta_0$ and the function

$$f(\xi, \nu) := \frac{1 - \xi^2 \left((1 + \kappa)(1 + \nu) + (1 + \kappa)(1 + 1/\nu)\Lambda_1^2\hat{\Lambda}_3 \right) - \kappa(1 + 1/\nu)\Lambda_1^2\Lambda_3}{1 + (1 + 1/\nu)\Lambda_1^2\Lambda_3}$$

is strictly smaller than $\theta_0 = \lim_{\nu \rightarrow \infty} f(0, \nu)$, there exists ν, ξ such that

$$\theta_A < f(\xi, \nu) < \theta_0.$$

Given κ_0 from Theorem 4.1 and assume $\kappa < \kappa_1 := \min \{ \kappa_0, \Lambda_1^{-2}\Lambda_3^{-1} \}$.

Case (A). Lemma 4.3.c and $\mu_\ell^2 \leq \kappa\eta_\ell^2$ prove that $\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)$ satisfies

$$\begin{aligned} & \left(1 - (1 + \kappa)\xi^2(1 + \nu) - (1 + \kappa)\xi^2(1 + 1/\nu)\Lambda_1^2\hat{\Lambda}_3 - \kappa(1 + 1/\nu)\Lambda_1^2\Lambda_3 \right) \eta_\ell^2 \\ & \leq (1 + (1 + 1/\nu)\Lambda_1^2\Lambda_3) \eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)). \end{aligned}$$

This reads $\theta_A \eta_\ell^2 \leq f(\xi, \nu) \eta_\ell^2 \leq \eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell))$ and implies that $\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)$ satisfies Dörfler marking in Case (A).

Let $\mathcal{M}_\ell =: \mathcal{M}_\ell^{(0)}$ be the set of marked elements in the Dörfler marking on level ℓ , while \mathcal{M}_ℓ^* is the optimal set of marked elements. Hence, there exists $0 < \Lambda_{\text{opt}} < \infty$ such that

$$|\mathcal{M}_\ell| \leq \Lambda_{\text{opt}} |\mathcal{M}_\ell^*| \leq \Lambda_{\text{opt}} |\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)|.$$

The control over $\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)$ of (A3) in Lemma 4.3.b results in

$$|\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)| \leq \Lambda_{\text{ref}} |\mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell| \leq \Lambda_{\text{ref}} \left(\frac{\Lambda_7 M}{\sqrt{1 - q\xi\sigma_\ell}} \right)^{1/s}.$$

Hence, $\Lambda_9 := \Lambda_{\text{opt}} \Lambda_{\text{ref}} \Lambda_7^{1/s} (\sqrt{1 - q\xi})^{-1/s}$ satisfies

$$|\mathcal{M}_\ell^{(0)}| = |\mathcal{M}_\ell| \leq \Lambda_9 M^{1/s} \sigma_\ell^{-1/s}. \quad (4.16)$$

Case (B). The output of `appx` with input triangulation \mathcal{T}_0 and input tolerance $\text{Tol} := \rho_B \mu_\ell^2$ on the level ℓ satisfies (B1). Since $\sigma_\ell^2 = \eta_\ell^2 + \mu_\ell^2 \leq (1 + 1/\kappa)\mu_\ell^2$ in Case (B), this leads to

$$|\mathcal{T}_{\text{Tol}}| - |\mathcal{T}_0| \leq \Lambda_5 (1 + 1/\kappa) \rho_B^{-1/(2s)} \sigma_\ell^{-1/s}.$$

According to [CR11, Rab15] for $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell \oplus \mathcal{T}_{\text{Tol}}$ there exists a finite sequence $(\mathcal{M}_\ell^{(k)})_{k=0, \dots, K(\ell)}$ of sets of marked element domains that $\mathcal{T}_\ell^{(0)} := \mathcal{T}_\ell$ and satisfies

$$\mathcal{T}_\ell^{(k+1)} = \text{REFINE}(\mathcal{T}_\ell^{(k)}, \mathcal{M}_\ell^{(k)}) \quad \text{for all } k = 0, \dots, K(\ell) - 1$$

leads to $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell^{(K(\ell))}$. This observation and the estimate for the overlay with the sequence $(\mathcal{M}_\ell^{(k)})_{k=0, \dots, K(\ell)}$ [CR11, Theorem 3.3] show

$$\sum_{k=0}^{K(\ell)} |\mathcal{M}_\ell^{(k)}| \leq |\mathcal{T}_{\text{Tol}}| - |\mathcal{T}_0| \lesssim \Lambda_5 (1 + 1/\kappa) \rho_B^{-1/(2s)} \sigma_\ell^{-1/s}. \quad (4.17)$$

The estimate from [CR11, Theorem 3.3] is for 2D only, however it is expected to hold in general.

Finish of the proof of “ \lesssim ”. It is proven in [CR11, Rab15] that the overhead control of [BDdV04, Ste08] holds in the sense that

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \leq \Lambda_{\text{BDdV}} \sum_{j=0}^{\ell-1} \sum_{k=0}^{K(j)} |\mathcal{M}_j^{(k)}|. \quad (4.18)$$

With (4.16)-(4.17) and Theorem 4.2.c, this proves

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \lesssim (\Lambda_5 + M^{1/s}) \sigma_\ell^{-1/s}. \quad (4.19)$$

Finally, $1 \leq |\mathcal{T}_\ell| - |\mathcal{T}_0|$ implies $1 + |\mathcal{T}_\ell| - |\mathcal{T}_0| \leq 2(|\mathcal{T}_\ell| - |\mathcal{T}_0|)$ while $|\mathcal{T}_\ell| = |\mathcal{T}_0|$ implies $1 \leq \sigma_\ell^{-1/s} (\Lambda_5 + M^{1/s})$. Hence (4.19) proves $\sigma_\ell(1 + |\mathcal{T}_\ell| - |\mathcal{T}_0|)^s \lesssim \Lambda_5^s + M$ and so “ \lesssim ” in the assertion of Theorem 2.1. \square

Proof of “ \gtrsim ” in (2.3) of Theorem 2.1.b. Given $N \in \mathbb{N}_0$ suppose that $\min \sigma(\mathbb{T}(N))$ is positive and so $\sigma_\ell > 0$ for all $\ell \in \mathbb{N}_0$ with $N_\ell := |\mathcal{T}_\ell| - |\mathcal{T}_0| \leq N$. This leads on the level ℓ in SAFEM to $N_{\ell+1} > N_\ell$ for it only stops with $\mathcal{T}_\ell = \mathcal{T}_{\ell+1} = \mathcal{T}_{\ell+2} = \dots$ when $\sigma_\ell = 0$. Hence there exists some level ℓ with $N_\ell < N \leq N_{\ell+1}$. This implies

$$(N+1)^s \min \sigma(\mathbb{T}(N)) \leq (N_{\ell+1} + 1)^s \sigma_\ell, \quad (4.20)$$

which is evident in case $\min \sigma(\mathbb{T}(N)) = 0$.

In Case (A) on the level ℓ of SAFEM, there is a one-level refinement to create $\mathcal{T}_{\ell+1}$ (indicated in Figure 2.1 for 2D), where each simplex in \mathcal{T}_ℓ creates a finite number $\leq K(n)$ of children in a completion step. The constant $K(n) \geq 2$ depends only on the spatial dimension n [GSS14]. This leads to the bound $|\mathcal{T}_{\ell+1}| \leq K(n) |\mathcal{T}_\ell|$ and then to

$$(N_{\ell+1} + 1)/(N_\ell + 1) \leq K(n) + (K(n) - 1)(|\mathcal{T}_0| - 1) \lesssim 1.$$

In Case (B) on the level ℓ of SAFEM, the refinement $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \mathcal{T}_{\text{Tot}}$ is controlled by $|\mathcal{T}_{\text{Tot}}| - |\mathcal{T}_0| \leq \Lambda_5 \text{Tot}^{-1/(2s)} \leq \Lambda_5 \rho_B^{-1/(2s)} \mu_\ell^{-1/s}$. Since $\sigma_\ell^2 \leq (1 + 1/\kappa) \mu_\ell^2$ in Case (B), the overlay estimate of [CKNS08, Ste07] proves

$$N_{\ell+1} - N_\ell \leq |\mathcal{T}_{\text{Tot}}| - |\mathcal{T}_0| \leq \Lambda_5 \rho_B^{-1/(2s)} (1 + 1/\kappa)^{1/(2s)} \sigma_\ell^{-1/s}.$$

This leads to the bound

$$2^{-s} (N_{\ell+1} + 1)^s \leq (N_\ell + 1)^s + \rho_B^{-1/2} (1 + 1/\kappa)^{1/2} \Lambda_5.$$

Consequently, in each of the Cases (A) and (B), it follows

$$(N_{\ell+1} + 1)^s \sigma_\ell \leq (K(n) + (K(n) - 1)(|\mathcal{T}_0| - 1))^s (N_\ell + 1)^s \sigma_\ell + 2^s \rho_B^{-s/2} (1 + 1/\kappa)^{s/2} \Lambda_5^s.$$

With $S := \sup_{\ell \in \mathbb{N}_0} (N_\ell + 1)^s \sigma_\ell$, this and (4.20) imply

$$(N+1)^s \min \sigma(\mathbb{T}(N)) \leq (K(n) + (K(n) - 1)(|\mathcal{T}_0| - 1))^s S + 2^s \rho_B^{-s/2} (1 + 1/\kappa)^{s/2} \Lambda_5^s.$$

Since this holds for any $N \in \mathbb{N}_0$, the previous N -independent upper bound is greater than or equal to the supremum M as well. This concludes the proof of “ \gtrsim ” in (2.3). \square

5. Application to mixed FEM. The a posteriori error analysis of mixed finite element schemes [Car97, Alo96] was completed in [CPS15] with a reliable and efficient error control in $H(\operatorname{div}, \Omega) \times L^2(\Omega)$, which is the natural functional analytical framework for the dual formulation of a Poisson model problem.

Given the right-hand side $f \in L^2(\Omega)$, the dual formulation of the Laplace equation on a 2D polygonal bounded simply-connected Lipschitz domain Ω seeks $p \in H(\operatorname{div}, \Omega)$ and $u \in L^2(\Omega)$ with

$$\begin{aligned} a(p, q) + b(q, u) &= 0 \quad \text{for all } q \in H(\operatorname{div}, \Omega), \\ b(p, v) &= -F(v) := -\int_{\Omega} f v \, dx \quad \text{for all } v \in L^2(\Omega). \end{aligned} \quad (5.1)$$

Therein, the bilinear forms model the L^2 scalar product and the divergence term,

$$a(p, q) := \int_{\Omega} p \cdot q \, dx \quad \text{and} \quad b(q, v) := \int_{\Omega} v \operatorname{div} q \, dx. \quad (5.2)$$

It is well established that the weak solution $u \in V := H_0^1(\Omega)$ to $-\Delta u = f$ in Ω specifies the flux $p := \nabla u$; the two formulations are equivalent and allow for unique solutions.

Given an admissible triangulation $\mathcal{T} \in \mathbb{T}$ let $(p_{RT}, u_{RT}) \in RT_0(\mathcal{T}) \times P_0(\mathcal{T})$ solve the discrete problem

$$\begin{aligned} a(p_{RT}, q_{RT}) + b(q_{RT}, u_{RT}) &= 0 \quad \text{for all } q_{RT} \in RT_0(\mathcal{T}), \\ b(p_{RT}, v_{RT}) &= -F(v_{RT}) \quad \text{for all } v_{RT} \in P_0(\mathcal{T}). \end{aligned} \quad (5.3)$$

Given the unique discrete solution (p_{RT}, u_{RT}) (resp. $(\widehat{p_{RT}}, \widehat{u_{RT}})$) with respect to the triangulation $\mathcal{T} \in \mathbb{T}$ (resp. its refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$), the estimators of [CPS15] and the distance function read

$$\begin{aligned} \eta^2(\mathcal{T}, K) &:= |K| \|p_{RT}\|_{L^2(K)}^2 + |K|^{1/2} \sum_{E \in \mathcal{E}(K)} \|[p_{RT}]_E \cdot \tau_E\|_{L^2(E)}^2, \\ \mu^2(K) &:= \|f - f_K\|_{L^2(K)}^2 \quad \text{for any } K \in \mathcal{T}, \\ \delta^2(\mathcal{T}, \hat{\mathcal{T}}) &:= \|\widehat{p_{RT}} - p_{RT}\|_{H(\operatorname{div}, \Omega)}^2. \end{aligned}$$

The standard 2D notation applies to the triangle K of area $|K|$ and its set $\mathcal{E}(K)$ of the three edges and the integral mean $f_K := \int f(x) \, dx / |K|$ of f . The jump $[\bullet]_E$ across an interior edge $E = \partial T_+ \cap \partial T_-$ with tangential normal vector τ_E and normal ν_E is the difference of the respective traces $[q]_E := q|_{T_+} - q|_{T_-}$ on E from the two neighboring triangles T_{\pm} . Homogeneous Dirichlet boundary data translate into homogeneous jumps on the boundary: $[q]_E := q|_{T_+}$ for $E \subset \partial\Omega$ with neighboring triangle T_+ .

It is remarkable that, in the lowest-order case at hand, the Lagrange multiplier u_{RT} does *not* enter the estimators and hence the distance function acts on the flux approximations only.

THEOREM 5.1 ((A1)-(A4)). *The estimators and distance functions satisfy (A1)-(A4) and (B2) for $\mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) := \mathcal{T} \setminus \hat{\mathcal{T}}$, $\Lambda_{\text{ref}} = 1 = \Lambda_6$, and $\hat{\Lambda}_3 = 0$.*

The estimator is reliable and efficient [CPS15] in that the exact (resp. discrete) solution (p, u) (resp. (p_{RT}, u_{RT}) with respect to $\mathcal{T} \in \mathbb{T}$) satisfies

$$\sigma(\mathcal{T}) \approx \|p - p_{RT}\|_{H(\operatorname{div}, \Omega)} + \|u - u_{RT}\|_{L^2(\Omega)}.$$

Hence the optimal rates of the estimators is equivalent to the optimal rates of the errors in terms of nonlinear approximation classes with respect to the natural norms in $H(\text{div}) \times L^2$ of the mixed FEM.

Proof of Theorem 5.1. It is straightforward to see that the estimators and distance function satisfy (A1)-(A2) with $\rho_2 := 2^{-1/4}$ and $\Lambda_1 = \Lambda_2 \approx 1$ stemming from trace and inverse estimates.

The proof of (A3) requires an intermediate solution $\widehat{p_{RT}}^* \in RT_0(\hat{\mathcal{T}})$ with respect to the fine triangulation $\hat{\mathcal{T}}$ to the above Poisson model problem with a piecewise constant right-hand side $\Pi_0 f \in P_0(\mathcal{T})$ with respect to the coarse triangulation \mathcal{T} . Let $\mathcal{E}' \subseteq \mathcal{E}$ be the subset of all edges such that at least one of the neighboring triangles $K \in \mathcal{T} \setminus \hat{\mathcal{T}}$ with $E \in \mathcal{E}(K)$ is refined ($K \notin \hat{\mathcal{T}}$). The divergence-free Raviart-Thomas function $\widehat{p_{RT}}^* - p_{RT}$ equals the rotated gradient of some continuous and piecewise affine function and so gives rise to a stability result

$$\|\widehat{p_{RT}}^* - p_{RT}\|_{L^2(\Omega)}^2 \lesssim \sum_{E \in \mathcal{E}'} |E| \|[p_{RT}]_E \cdot \tau_E\|_{L^2(E)}^2 \quad (5.4)$$

proved via a discrete Helmholtz decomposition (cf. e.g. [CHX09, Thm 5.6] for references and the arguments) for a simply connected domain Ω .

The discrete inf-sup condition (with respect to the finer mesh $\hat{\mathcal{T}}$) leads to some $\widehat{q_{RT}} \in RT_0(\hat{\mathcal{T}})$ and $\widehat{v}_0 \in P_0(\hat{\mathcal{T}})$ with norm $\|\widehat{q_{RT}}\|_{H(\text{div}, \Omega)} + \|\widehat{v}_0\|_{L^2(\Omega)} \lesssim 1$ and

$$\begin{aligned} \text{LHS} &:= \|\widehat{p_{RT}} - p_{RT}\|_{H(\text{div}, \Omega)} + \|\widehat{u_{RT}} - u_{RT}\|_{L^2(\Omega)} \\ &= a(\widehat{p_{RT}} - p_{RT}, \widehat{q_{RT}}) + b(\widehat{q_{RT}}, \widehat{u_{RT}} - u_{RT}) + b(\widehat{p_{RT}} - p_{RT}, \widehat{v}_0). \end{aligned} \quad (5.5)$$

The discrete equations (5.3) on the fine level $\hat{\mathcal{T}}$ and $\text{div } p_{RT} = -\Pi_0 f$ show

$$\text{LHS} = -a(p_{RT}, \widehat{q_{RT}}) - b(\widehat{q_{RT}}, u_{RT}) - F(\widehat{v}_0 - \Pi_0 \widehat{v}_0). \quad (5.6)$$

Given $\widehat{q_{RT}}$ with bounded norm, let q_{RT} denote the mixed finite element solution to a Poisson model problem with right-hand side $-\Pi_0 \text{div } \widehat{q_{RT}} \in P_0(\mathcal{T})$. This leads to $\|q_{RT}\|_{H(\text{div}, \Omega)} \lesssim 1$ and

$$b(\widehat{q_{RT}}, u_{RT}) = b(q_{RT}, u_{RT}) = -a(p_{RT}, q_{RT}).$$

With $\|\widehat{v}_0\| \lesssim 1$, the combination of the two previously displayed formulas shows

$$\text{LHS} \lesssim \|\Pi_0 f - \Pi_0 f\|_{L^2(\Omega)} + a(p_{RT}, q_{RT} - \widehat{q_{RT}}).$$

The Cauchy-Schwarz inequality leads to

$$a(p_{RT}, q_{RT} - \widehat{q_{RT}}) = a(p_{RT} - \widehat{p_{RT}}^*, q_{RT} - \widehat{q_{RT}}) + a(\widehat{p_{RT}}^*, q_{RT} - \widehat{q_{RT}}) \quad (5.7)$$

$$\lesssim \|p_{RT} - \widehat{p_{RT}}^*\|_{L^2(\Omega)} + a(\widehat{p_{RT}}^*, q_{RT} - \widehat{q_{RT}}). \quad (5.8)$$

Due to (5.4) it remains to analyze the latter term. The test function equals [Mar85, AB85]

$$q_{RT} - \widehat{q_{RT}} = \nabla_{NC} \widehat{v_{CR}} + \text{curl } \widehat{\beta}_C + 1/2 ((\Pi_0 - 1) \text{div } \widehat{q_{RT}}) (\bullet - \text{mid}(\hat{\mathcal{T}})) \quad (5.9)$$

for unique discrete functions $\widehat{v_{CR}} \in \text{CR}_0^1(\hat{\mathcal{T}})$ and $\widehat{\beta}_C \in S^1(\hat{\mathcal{T}})/\mathbb{R}$ on the fine level, all bounded by the left-hand side $\lesssim 1$. The same argument shows

$$\widehat{p_{RT}}^* = \nabla_{NC} \widehat{u_{CR}}^* - 1/2 (\Pi_0 f) (\bullet - \text{mid}(\hat{\mathcal{T}})) \quad (5.10)$$

for some $\widehat{u_{CR}}^* \in CR_0^1(\hat{\mathcal{T}})$. The remaining term $a(\widehat{p_{RT}}^*, q_{RT} - \widehat{q_{RT}})$ equals

$$\int_{\Omega} \widehat{p_{RT}}^* \cdot \nabla_{NC} \widehat{v_{CR}} dx + \frac{1}{2} \int_{\Omega} \widehat{p_{RT}}^* \cdot (x - \text{mid}(\hat{\mathcal{T}}))(\Pi_0 - 1) \text{div} \widehat{q_{RT}} dx.$$

This, the representation (5.9) of $q_{RT} - \widehat{q_{RT}}$, and an integration by parts show

$$\begin{aligned} \int_{\Omega} \widehat{p_{RT}}^* \cdot \nabla_{NC} \widehat{v_{CR}} dx &= \int_{\Omega} \nabla_{NC} \widehat{u_{CR}}^* \cdot \nabla_{NC} \widehat{v_{CR}} dx \\ &= \int_{\Omega} \nabla_{NC} \widehat{u_{CR}}^* \cdot (q_{RT} - \widehat{q_{RT}}) dx = \int_{\Omega'} \widehat{u_{CR}}^* \text{div}(\widehat{q_{RT}} - q_{RT}) dx. \end{aligned}$$

Therein, Ω' is the interior of the $\bigcup(\mathcal{T} \setminus \hat{\mathcal{T}})$, the union of the elements in $\mathcal{T} \setminus \hat{\mathcal{T}}$. Since $\text{div}(\widehat{q_{RT}} - q_{RT}) = (1 - \Pi_0) \text{div} \widehat{q_{RT}}$ is L^2 perpendicular to $P_0(\mathcal{T})$ (and so vanishes on $\mathcal{T} \cap \hat{\mathcal{T}}$ outside of Ω'), a discrete Poincaré inequality proves that this is bounded from above by $\lesssim \|h_{\mathcal{T}} \nabla_{NC} \widehat{u_{CR}}^*\|_{L^2(\Omega')}$. Since $(1 - \Pi_0) \text{div} \widehat{q_{RT}}$ vanishes outside of Ω' and has a bounded L^2 norm, the second integral reads

$$\frac{1}{2} \int_{\Omega} \widehat{p_{RT}}^* \cdot (x - \text{mid}(\hat{\mathcal{T}}))(\Pi_0 - 1) \text{div} \widehat{q_{RT}} dx \lesssim \|h_{\mathcal{T}} \widehat{p_{RT}}^*\|_{L^2(\Omega')}.$$

The combination of the three previously displayed formulas and a triangle inequality lead to

$$\begin{aligned} a(\widehat{p_{RT}}^*, q_{RT} - \widehat{q_{RT}}) &\lesssim \|h_{\mathcal{T}} \widehat{p_{RT}}^*\|_{L^2(\Omega')} + \|h_{\mathcal{T}} \nabla_{NC} \widehat{u_{CR}}^*\|_{L^2(\Omega')} \\ &\lesssim \|h_{\mathcal{T}} \widehat{p_{RT}}^*\|_{L^2(\Omega')} + \|h_{\mathcal{T}}(\widehat{p_{RT}}^* - \nabla_{NC} \widehat{u_{CR}}^*)\|_{L^2(\Omega')}. \end{aligned}$$

The representation (5.10) shows that the last term is equal to

$$1/2 \|h_{\mathcal{T}}(\Pi_0 f)(\bullet - \text{mid}(\hat{\mathcal{T}}))\|_{L^2(\Omega')} \lesssim \|h_{\mathcal{T}}^2 \text{div} p_{RT}\|_{L^2(\Omega')}.$$

An inverse estimate for p_{RT} on any $K \in \mathcal{T} \setminus \hat{\mathcal{T}}$ leads to

$$\|h_{\mathcal{T}}^2 \text{div} p_{RT}\|_{L^2(\Omega')} \lesssim \|h_{\mathcal{T}} p_{RT}\|_{L^2(\Omega')}.$$

A triangle inequality plus $\|h_{\mathcal{T}}\|_{L^\infty(\Omega')} \lesssim 1$ prove

$$\|h_{\mathcal{T}} \widehat{p_{RT}}^*\|_{L^2(\Omega')} \lesssim \|h_{\mathcal{T}} p_{RT}\|_{L^2(\Omega')} + \|\widehat{p_{RT}}^* - p_{RT}\|_{L^2(\Omega')}.$$

The combination of the above estimates (i.e. (5.6), (5.7) and the three previously displayed formulas) shows that

$$\begin{aligned} \|\widehat{p_{RT}} - p_{RT}\|_{H(\text{div}, \Omega)} + \|\widehat{u_{RT}} - u_{RT}\|_{L^2(\Omega)} \\ \lesssim \|\widehat{p_{RT}}^* - p_{RT}\|_{L^2(\Omega)} + \|h_{\mathcal{T}} p_{RT}\|_{L^2(\Omega')} + \|\widehat{\Pi_0} f - \Pi_0 f\|_{L^2(\Omega)}. \end{aligned} \tag{5.11}$$

The L^2 orthogonal projection Π_0 (resp. $\widehat{\Pi_0}$) with respect to $\mathcal{T} \in \mathbb{T}$ (resp. its refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$) leads to the data approximation term

$$\|\widehat{\Pi_0} f - \Pi_0 f\|_{L^2(\Omega)}^2 = \mu^2(\mathcal{T}) - \mu^2(\hat{\mathcal{T}}).$$

The combination of this with (5.4) and (5.11) proves (A3) in the sharper form

$$\delta^2(\mathcal{T}, \hat{\mathcal{T}}) \leq \|\widehat{p_{RT}} - p_{RT}\|_{H(\text{div}, \Omega)}^2 + \|\widehat{u_{RT}} - u_{RT}\|_{L^2(\Omega)}^2 \lesssim \eta^2(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}}) + \mu^2(\mathcal{T}) - \mu^2(\hat{\mathcal{T}}).$$

The proof of (A4) recalls the L^2 quasiorthogonality of the flux errors of [CHX09, Thm 3.2] or [CR11, Lemma 4.3 and (4.4)] in the form

$$\|p_{\ell+1} - p_\ell\|_{L^2(\Omega)}^2 + \|p - p_{\ell+1}\|_{L^2(\Omega)}^2 - \|p - p_\ell\|_{L^2(\Omega)}^2 \lesssim \|p - p_{\ell+1}\|_{L^2(\Omega)} \operatorname{osc}(f_{\ell+1}, \mathcal{T}_\ell).$$

The mixed FEM fixes the divergence of the flux approximations, $-\operatorname{div} p_\ell = \Pi_\ell f =: f_\ell$, and their orthogonality

$$\|f_{\ell+1} - f_\ell\|_{L^2(\Omega)}^2 + \|f - f_{\ell+1}\|_{L^2(\Omega)}^2 - \|f - f_\ell\|_{L^2(\Omega)}^2 = 0$$

leads (for all $\ell \in \mathbb{N}$) in the aforementioned L^2 quasiorthogonality to

$$\|p_{\ell+1} - p_\ell\|_{H(\operatorname{div}, \Omega)}^2 + \|p - p_{\ell+1}\|_{H(\operatorname{div}, \Omega)}^2 - \|p - p_\ell\|_{H(\operatorname{div}, \Omega)}^2 \lesssim \|p - p_{\ell+1}\|_{L^2(\Omega)} \operatorname{osc}(f_{\ell+1}, \mathcal{T}_\ell).$$

For any $0 < \varepsilon$ with $\varepsilon \Lambda_3 < 1$ and the multiplicative constant $C \approx 1$ hidden in the notation \lesssim the sum of those estimates results for any $\ell, m \in \mathbb{N}_0$ in

$$\begin{aligned} \sum_{k=\ell}^{\ell+m} \|p_{k+1} - p_k\|_{H(\operatorname{div}, \Omega)}^2 &\leq \|p - p_\ell\|_{H(\operatorname{div}, \Omega)}^2 + \varepsilon / \Lambda_3 \sum_{k=\ell}^{\ell+m-1} \|p - p_{k+1}\|_{L^2(\Omega)}^2 \\ &\quad + C^2 \Lambda_3 / \varepsilon \sum_{k=\ell}^{\ell+m} \operatorname{osc}^2(f_{k+1}, \mathcal{T}_k). \end{aligned} \quad (5.12)$$

For a sequence of uniformly refined meshes $\hat{\mathcal{T}}$, the discrete reliability (A3) leads to the reliability of [CPS15],

$$\|p - p_\ell\|_{H(\operatorname{div}, \Omega)}^2 \leq \Lambda_3 \eta_\ell^2 := \eta^2(\mathcal{T}_\ell) \quad \text{for all } \ell \in \mathbb{N}_0.$$

The oscillation $\operatorname{osc}(f_{k+1}, \mathcal{T}_k) = \|h_\ell(f_{k+1} - f_k)\|_{L^2(\Omega)}$ is bounded by $\|h_\ell\|_{L^\infty(\Omega)} \|f_{k+1} - f_k\|_{L^2(\Omega)}$. With $h_{\max} := \|h_0\|_{L^\infty(\Omega)} \lesssim 1$, the L^2 orthogonality of the integrands shows

$$\sum_{k=\ell}^{\ell+m} \operatorname{osc}^2(f_{k+1}, \mathcal{T}_k) \leq h_{\max} \|f_{\ell+m+1} - f_\ell\|_{L^2(\Omega)}^2 \leq h_{\max} \|f - f_\ell\|_{L^2(\Omega)}^2.$$

The combination of the previous estimates with (5.12) leads to the quasiorthogonality (A4) in the form

$$\sum_{k=\ell}^{\ell+m} \delta_{k,k+1}^2 \leq \Lambda_3 \eta_\ell^2 + \varepsilon \sum_{k=\ell+1}^{\ell+m} \eta_k^2 + C \Lambda_3 h_{\max} / \varepsilon \mu^2(\mathcal{T}_\ell).$$

This is $(A4_\varepsilon)$ with $\Lambda_{4(\varepsilon)} := \max\{\Lambda_3, C \Lambda_3 h_{\max} / \varepsilon\}$ for any $\varepsilon > 0$. This and (A1)-(A2) imply (A4) owing to Theorem 3.1. The remaining details are omitted. \square

6. Application to least-squares FEM. The div least-squares formulation [BG09] of the Poisson model example of the previous section seeks the minimizer (p, u) of the functional

$$\operatorname{LS}(f; q, v) := \|f + \operatorname{div} q\|_{L^2(\Omega)}^2 + \|q - \nabla v\|_{L^2(\Omega)}^2$$

amongst $(q, v) \in H(\operatorname{div}, \Omega) \times H_0^1(\Omega)$. The functional $\operatorname{LS}(f; \bullet)$ is indeed a natural a posteriori error estimator. Given any admissible triangulation $\mathcal{T} \in \mathbb{T}$, the least-squares FEM seeks the minimizer $(p_{\text{LS}}, u_{\text{LS}})$ of $\operatorname{LS}(f; \bullet)$ in the discrete subspace

$RT_0(\mathcal{T}) \times S_0^1(\mathcal{T})$. This leads in [CP15] to the alternative a posteriori error estimate with

$$\tilde{\eta}^2(\mathcal{T}, K) := \|(1 - \Pi_0)p_{LS}\|_{L^2(K)}^2 + |K|^{1/2} \sum_{E \in \mathcal{E}(K)} \|[p_{LS}]_E \cdot \tau_E\|_{L^2(E)}^2 \quad (6.1)$$

$$+ |K|^{1/2} \sum_{E \in \mathcal{E}(K) \setminus \mathcal{E}(\partial\Omega)} \|\partial u_{LS}/\partial \nu_E\|_{L^2(E)}^2,$$

$$\mu^2(K) := \|f - \Pi_0 f\|_{L^2(K)}^2 \quad \text{for any } K \in \mathcal{T}. \quad (6.2)$$

Given a refined triangulation $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ with discrete solutions $(\widehat{p_{LS}}, \widehat{u_{LS}})$, the distance

$$\delta^2(\mathcal{T}, \hat{\mathcal{T}}) := \text{LS}(f; p_{LS}, u_{LS}) - \text{LS}(f; \widehat{p_{LS}}, \widehat{u_{LS}}) = \text{LS}(0; \widehat{p_{LS}} - p_{LS}, \widehat{u_{LS}} - u_{LS})$$

is equivalent to the norm of the difference $(\widehat{p_{LS}} - p_{LS}, \widehat{u_{LS}} - u_{LS})$ of the two discrete solutions in $H(\text{div}, \Omega) \times H_0^1(\Omega)$ [BG09].

THEOREM 6.1 (A1–A4). *The estimators and distance function satisfy (A1)–(A4) and (B2) for $\mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) := \mathcal{T} \setminus \hat{\mathcal{T}}$, $\Lambda_{\text{ref}} = 1 = \Lambda_6$, and (QM).*

Since the estimator is reliable and efficient, for the discrete solution (p_{LS}, u_{LS}) with respect to $\mathcal{T} \in \mathbb{T}$,

$$\sigma(\mathcal{T}) \approx \|p - p_{LS}\|_{H(\text{div}, \Omega)} + \|u - u_{LS}\|_{H^1(\Omega)},$$

the optimal rates of the estimators is equivalent to the optimal rates of the errors in terms of nonlinear approximation classes with respect to the natural norms in $H(\text{div}) \times H^1$ of the least-squares FEM.

Proof of Theorem 6.1. The proofs are essentially contained in [CP15]. The axioms (A1)–(A2) are standard and (A3) follows from

$$\text{LS}(f; p_{LS}, u_{LS}) \lesssim \eta^2(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}}) + \mu^2(\mathcal{T}) + \text{LS}(f; \widehat{p_{LS}}, \widehat{u_{LS}})$$

(this is [CP15, p 59, line 24] in different notation) and another reliability estimate $\text{LS}(f; \widehat{p_{LS}}, \widehat{u_{LS}}) \approx \sigma^2(\hat{\mathcal{T}})$ from [CP15, Thm 3.1]. Notice that $\hat{\Lambda}_3$ does *not* need to be small (at least for coarse meshes according to the Remark 2) and hence Theorem 3.2 cannot be applied to ensure (QM) in general. On the other hand, any conforming discretization reduces the least-squares functions and so

$$\sigma^2(\hat{\mathcal{T}}) \approx \text{LS}(f; \widehat{p_{LS}}, \widehat{u_{LS}}) \leq \text{LS}(f; p_{LS}, u_{LS}) \approx \sigma^2(\mathcal{T})$$

immediately leads to (QM). The same argument plus the reliability of [CP15, Theorem 3.1] prove (A4) even in the sharper form of an orthogonality. The remaining details are omitted. \square

REMARK 2. *A detailed analysis of [CP15] (beyond this paper) with reduced elliptic regularity suggests that $\hat{\Lambda}_3 \leq C(\epsilon) h_{\max}^{1/2+\epsilon}$ for small $\epsilon > 0$ (depending on the interior angles of the domain) and some constant $C(\epsilon)$. Hence $\hat{\Lambda}_3$ tends to zero as the maximal mesh-size h_{\max} tends to zero and so Theorem 3.2 is applicable for sufficiently fine meshes.*

REMARK 3. *The analysis also allows optimal convergence rates for modified estimators such as*

$$\eta^2(K) := |K| \|\text{D} p_{LS}\|_{L^2(K)} + |K|^{1/2} \|[p_{LS} - \nabla u_{LS}]_{\partial K}\|_{L^2(\partial K)}$$

with $[p_{LS} - \nabla u_{LS}]_{\partial K} := (p_{LS} - \nabla u_{LS})|_K$ along $E \in \mathcal{E}(\partial\Omega)$ with $K = \bar{\omega}_E$. This estimator is close to the least-squares functional estimators, but not equivalent.

REFERENCES

- [AB85] D. N. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *RAIRO, Mathematical Modelling and Numerical Analysis*, 19(1):7–32, 1985.
- [Alo96] A. Alonso. Error estimators for a mixed method. *Numer. Math.*, 74:385–395, 1996.
- [BC] P. Bringmann and C. Carstensen. An adaptive least-squares FEM for the Stokes equations with optimal convergence rates. Submitted to *Numer. Math.*, 2015.
- [BCS] P. Bringmann, C. Carstensen, and G. Starke. An adaptive least-squares FEM for linear elasticity with optimal convergence rates. In preparation.
- [BDdV04] P. Binev, W. Dahmen, and R. de Vore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97:219–268, 2004.
- [BdV04] P. Binev and R. de Vore. Fast computation in adaptive tree approximation. *Numer. Math.*, 97:193–217, 2004.
- [BG09] Pavel B. Bochev and Max D. Gunzburger. *Least-Squares Finite Element Methods*. Springer, 2009.
- [BM08] R. Becker and S. Mao. An optimally convergent adaptive mixed finite element method. *Numer. Math.*, 111:35–54, 2008.
- [Car97] C. Carstensen. A posteriori error estimate for the mixed finite element method. *Math. Comp.*, 66:465–476, 1997.
- [CFPP14] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. Axioms of adaptivity. *Comput. Methods Appl. Math.*, 67(6):1195–1253, 2014.
- [CHX09] L. Chen, M. Holst, and J. Xu. Convergence and optimality of adaptive mixed finite element methods. *Math. Comp.*, 78(265):35–53, 2009.
- [CKNS08] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.
- [CP15] C. Carstensen and E.-J. Park. Convergence and optimality of adaptive least squares finite element methods. *SIAM J. Numer. Anal.*, 53:43–62, 2015.
- [CPS15] C. Carstensen, D. Peterseim, and A. Schröder. The norm of a discretized gradient in $H(\operatorname{div})^*$ for a posteriori finite element error analysis. *Numer. Math.*, 2015.
- [CR11] C. Carstensen and H. Rabus. An optimal adaptive mixed finite element method. *Math. Comp.*, 80(274):649–667, 2011.
- [GSS14] Dietmar Gallistl, Mira Schedensack, and Rob Stevenson. A remark on newest vertex bisection in any space dimension. *Computational methods in applied mathematics*, 14(3), 2014.
- [Mar85] L. D. Marini. An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method. *SIAM J. Numer. Anal.*, 22:493–496, 1985.
- [Rab15] H. Rabus. Quasi-optimal convergence of AFEM based on separate marking – Part I and II. *Journal of Numerical Mathematics*, 23(2):137–156, 157–174, 2015.
- [Ste07] R. Stevenson. Optimality of a standard adaptive finite element method. *Foundations of Computational Mathematics*, 7(2):245–269, 2007.
- [Ste08] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77:227–241, 2008.